

# Generation of OPLS-like Charges from Molecular Electrostatic Potential Using Restraints

RICHARD H. HENCHMAN, JONATHAN W. ESSEX

*Department of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom*

*Received 26 March 1998; accepted 6 November 1998*

**ABSTRACT:** A new method for generating atom-centered charges for use in condensed phase computer simulations is presented, which is based on a restrained electrostatic potential (RESP) procedure. Charges are calculated from a least-squares fit to the quantum mechanical electrostatic potential with a restraint applied to reduce their magnitude. The restraint developed here offers advantages over that used in RESP. The magnitude of the restraint is optimized to yield charges as close to the equivalent OPLS values as possible while still reproducing the molecule's electrostatic potential. A cross-validation analysis is used to show that the restraint is insensitive to the selection of OPLS molecules from which it is derived. Thus, with this method, OPLS-like charges may be produced from the electrostatic potential for atom types not in the OPLS force field. In addition, the restraint is shown to reduce the conformational dependence of the charges. © 1999 John Wiley & Sons, Inc. *J Comput Chem* 20: 483–498, 1999

**Keywords:** electrostatic potential; charges; restraint; OPLS; parameters

## Introduction

The simplest, most widespread method of representing the electrostatic interaction between molecules in a computer simulation is to approximate it with a Coulombic potential between sets of atom-centered charges. The deficiencies associated

with this method include the monopole approximation itself,<sup>1–3</sup> the absence of explicit polarization in their parameterization,<sup>4</sup> and the dependence of charges on the geometry of the molecule from which they were obtained and its conformation in particular.<sup>5–11</sup> Although extensions of the basic method to overcome these problems are available, the use of charges is extremely common in the simulation of molecular systems and they are almost ubiquitous for biological molecules. This

Correspondence to: J. W. Essex; e-mail: jwe1@soton.ac.uk

is in part historical; but it is also largely due to the computational expense resulting from the increased complexity, implementation, and reparameterization required to incorporate the extensions. Even if these extensions cannot be directly introduced, including an implicit treatment of them in the charge deviation method is desirable.

The use of charges in a computer simulation is an *approximate* method for modeling electrostatic interactions; consequently, there is neither a single set of parameters capable of reproducing every piece of data nor is there a unique, correct way to derive these parameters. Indeed, there are a wide variety of methods to parameterize charges. Charges can be derived from either experimental or theoretical data. Of the methods using experimental information, charges may be obtained from the electron densities derived from X-ray diffraction studies.<sup>12</sup> Alternatively, the charges may be optimized to reproduce structural and thermodynamic experimental data for pure liquids or aqueous solutions of a molecule. This procedure is used in the popular OPLS force field.<sup>13</sup> Hybrid techniques also exist that use experimental data and quantum mechanical calculations such as the charge equilibration method,<sup>14</sup> while the atomic polar tensors method can use either.<sup>15–17</sup> There is a large collection of purely quantum mechanical methods that partition the electron population between each atom according to the calculated molecular orbitals<sup>18–21</sup> as in, for example, the Mulliken population analysis.<sup>18</sup> However, for all of these, the partitioning is somewhat arbitrary and the calculated charges are basis set dependent and generally give a poor reproduction of the molecular electrostatic potential (MEP).<sup>22,23</sup> An alternative approach is to optimize charges to reproduce various quantum mechanical quantities such as the calculated interaction energy of a collection of molecular complexes<sup>24</sup> or the MEP,<sup>22,25–29</sup> the molecule's electric field,<sup>30</sup> or a distributed multipole potential.<sup>31–33</sup>

Because the OPLS method optimizes charges to reproduce condensed phase properties, they are arguably the parameters of choice for condensed phase simulations. In addition, the OPLS method is best able to address the problems previously discussed by including approximate, implicit treatments of polarization and conformational averaging. However, there are a number of limitations in the OPLS method. First, extensive computer simulations are needed. Second, for each chemical functionality, a simple molecule containing that functionality is required for which there are experi-

mental data either of the liquid phase or an aqueous solution. Third, the charges for a given functional group may not always be transferable, particularly for two adjacent polar groups. Thus, because OPLS charges are not available for all types and combinations of functional groups, another reliable and practical method must be used. Because charges are used to model intermolecular interactions, they should be optimized to reproduce molecular data in regions lying outside a molecule's van der Waals volume. Using the MEP to calculate charges is therefore a possibility, whereas use of the electron density is not. Furthermore, the monopole approximation becomes less severe at these longer ranges. Therefore, the widely used MEP procedure seems the next best and simplest alternative. The calculation of these electrostatic potential derived (EPD) charges is described in more detail in the Method section, but in essence the method works by calculating the charges that produce an EP that most closely matches the MEP at a number of points spaced evenly around the molecule. The procedure is easy to implement, is practical for even quite large molecules, and, most importantly, can be applied to any functionality.

Nevertheless, a number of difficulties still exist with the use of EPD charges. First, some degree of polarization is required to model that induced in the condensed phase. Typically, this is introduced in an *ad hoc* fashion through the use of a low-quality basis set such as 6–31G\*,<sup>34</sup> although it is possible to calculate charges in a polarizable continuum.<sup>35</sup> Second, the MEP generally does not provide enough information to determine statistically valid charges for all atoms in a molecule, which was demonstrated by a singular value decomposition analysis by Francl et al.<sup>28</sup> This essentially implies that the underdetermined atoms may adopt charges over a wide range of values without significantly affecting the charge EP (CEP) and thus the quality of the fit to the MEP. This is particularly the case for atoms with typically three or four neighbors, commonly termed "buried atoms," whose charges contribute less to the CEP in regions outside the molecule than surface atom charges. Also, when preserving overall molecular charge, the charge removed from buried atoms can be spread more evenly over their larger number of neighbors. Third, EPD charges often do not accord with chemical intuition. This manifests itself in a number of ways. EPD charges can be large in magnitude,<sup>36</sup> larger than chemical intuition would suggest, and also larger than the equivalent OPLS

parameters. This raises compatibility problems if some combination of the two force fields is desired. EPD charges can also vary significantly for chemically similar atoms in a given homologous series, raising difficulties if a transferable parameter set is desired. Furthermore, the polarity of EPD charges does not always correlate with electronegativities, as is often found for C–H bonds. However, this should not be unexpected given that EPD charges are derived from the MEP and not the charge density. Fourth, EPD charges often have quite significant conformational dependence,<sup>5–11</sup> a problem exacerbated by their large magnitude.

Given these difficulties with the EPD charges and the unavailability of OPLS charges for some functionalities, ideally a method is needed that has the speed and flexibility of the MEP procedure that generates electrostatic parameters of comparable magnitude to the OPLS force field. This article proposes such a method. This restrained EPD (REPD) charge method is an extension of the restrained electrostatic potential (RESP) procedure of Bayly et al.<sup>36</sup> and uses the statistical indeterminacy of buried charges to our advantage.

In the RESP method the charges  $q_j$  are fitted to the MEP while simultaneously being minimized. In this way the large size of EPD charges noted previously is reduced. This is accomplished using a restraining function of the following form in the least-squares fit

$$\chi_{\text{rest}}^2 = a \sum_{j=1}^M \left( (q_j^2 + b^2)^{1/2} - b \right), \quad (1)$$

where  $M$  is the number of non-hydrogen atoms. The parameter  $a$  determines the strength of the restraint and takes two values, one for the initial restraint and one in the averaging of conformationally equivalent atoms. These values are 0.0005 and 0.001, respectively. The parameter  $b$  determines the tightness of the hyperbola and is given a value of 0.1. The hyperbolic restraining function restrains all charges with approximately the same force except for those with a magnitude comparable to  $b$  for which the restraining force is smaller. On closer inspection, however, a number of deficiencies in the RESP method are apparent: the effect of the restraint is dependent on the selection of points used in the fitting procedure; the charges do not vary uniformly with  $a$ ; hydrogen atoms are selectively excluded from the restraint, and parameters  $a$  and  $b$  are somewhat arbitrary.

This article describes a new restraining function that addresses these deficiencies. The new restraint is independent of the number of points used in the fitting procedure, is quadratic in charge, is applied uniformly to all atoms, and contains one adjustable parameter that is optimized to yield charges in close agreement with the equivalent OPLS parameters for a wide range of molecules. A cross-validation analysis demonstrates that this parameter is largely insensitive to the choice of molecules used in parameterization. The method is shown to be applicable to the 6-31G\* and 6-31+G\* basis sets. Finally, the use of the restraint is shown to reduce the conformational dependence of REPD charges.

## REPD Charge Method

To calculate charges that reproduce an MEP, an appropriate molecular geometry and MEP are required. The MEP was calculated at a selection of points spaced evenly around the molecule at distances considered significant for interaction with other molecules. Charges were then calculated by performing a least-squares fit at these points between the MEP and the CEP with appropriate charge averaging of equivalent atoms. The effect of including a restraining function in the least-squares fit was examined.

*Ab initio* optimized geometries were calculated at the Hartree–Fock (HF) level with 6-31G\* and 6-31+G\* basis sets using Gaussian 94.<sup>37</sup> For simulations in which conformationally equivalent atoms are allowed to interconvert freely, such as the hydrogen atoms of a rotating methyl group, these atoms must have identical parameters. Therefore, bond lengths and angles were constrained to be the same and dihedrals set at appropriate values for these atoms. HF/6-31G\* geometries were used because they agree well with the experiment<sup>27,34</sup>; are easily calculated for small molecules; and, unlike experimental geometries, are readily obtainable for any common molecular functionality, the basis set permitting. The almost identical HF/6-31+G\* geometries have similar properties and are only moderately more computationally demanding to calculate.

The MEP may be calculated at any arbitrary point in space,  $r_i$ , from an *ab initio* calculation using the equation

$$V_i = \sum_j Z_j / r_{ij} - \sum_{\mu} \sum_{\nu} P_{\mu\nu} \int \frac{\chi_{\mu}(r) \chi_{\nu}(r)}{|r_i - r|} dr, \quad (2)$$

where  $r_{ij}$  is the distance between  $r_i$  and atom  $j$  with charge  $Z_j$  and  $P_{\mu\nu}$  is the density matrix element for the atomic basis functions  $\chi_\mu$  and  $\chi_\nu$ . The *ab initio* protocol used to calculate the MEP is identical to that used for the molecular geometry: either HF/6-31G\* or HF/6-31+G\*.

Points were selected according to the method of Singh and Kollman<sup>26</sup> in which points are spaced using Connolly's algorithm<sup>38</sup> on four spheres centered on each atom with radii 1.4, 1.6, 1.8, and 2.0 times the atom's van der Waals radius. Points at a distance less than 1.4 times the van der Waals radius of any atom were excluded. The point density on each sphere was 1 point Å<sup>-2</sup>. GAMESS<sup>39</sup> was used to generate a table containing the MEP at each point. It is well known that charges vary somewhat with point selection algorithms<sup>10,22,23,26,27,40,41</sup> and, in particular, the number of points, the distance of points from each atom, the molecular orientation, and, for grid-based algorithms, the molecular position within the grid. However, we show that the number of points used in this work is adequate to determine a charge set that well reproduces the MEP.

Charges ( $q_j$ ) were calculated by minimizing the difference between the CEP, given by

$$V'_i = \sum_{j=1}^M \frac{q_j}{r_{ij}}, \quad (3)$$

and the MEP.  $M$  is the number of atoms in the molecule. This difference may be quantified by the following  $\chi^2$  function,

$$\chi^2 = \sum_{i=1}^N (V_i - V'_i)^2, \quad (4)$$

where  $N$  is the number of fitting points. To find charges that minimize  $\chi^2$  in eq. (4), setting the derivative of  $\chi^2$  with respect to each charge to zero produces  $M$  equations linear in  $q_j$  of the form

$$\frac{\partial \chi^2}{\partial q_k} = -2 \sum_{i=1}^N \frac{1}{r_{ik}} \left( V_i - \sum_{j=1}^M \frac{q_j}{r_{ij}} \right) = 0, \quad (5)$$

which can be easily solved in matrix form for  $\mathbf{q}$ ,

$$\mathbf{A}\mathbf{q} = \mathbf{b}, \quad (6)$$

where the elements of  $\mathbf{A}$  are given by

$$A_{jk} = \sum_{i=1}^N \frac{1}{r_{ij}r_{ik}} \quad (7)$$

and the elements of  $\mathbf{b}$  by

$$b_k = \sum_{i=1}^N \frac{V_i}{r_{ik}}. \quad (8)$$

Two measures were taken to ensure a correct overall molecular charge. A total charge constraint, implemented using Lagrange multipliers,<sup>27</sup> is imposed, and subsequent rounding to three decimal places for use in a simulation usually requires charges to be manually readjusted. In this adjustment buried atom charges are given preference because they have less effect on the CEP. The quality of the fit can be assessed by the relative root mean square (RRMS) of  $\chi^2$ , which is defined by

$$\text{RRMS} = \left( \chi^2 / \sum_{i=1}^N V_i^2 \right)^{1/2}. \quad (9)$$

A restraining function that has the effect of reducing a charge's magnitude may be added to eq. (4). The restraining function developed in this work, which is quadratic in charge, is given by

$$\chi_{\text{rest}}^2 = a \sum_{j=1}^M \sum_{i=1}^N \left( \frac{q_j}{r_{ij}} \right)^2, \quad (10)$$

where  $a$  determines the strength of the restraint. The diagonal elements of the  $\mathbf{A}$  matrix in eq. (6) after the addition of the restraint now become

$$A_{kk} = \sum_{i=1}^N \frac{(1+a)}{r_{ik}^2}, \quad (11)$$

while the off-diagonal elements remain the same. Because the parameter  $a$  can be factored into the sum over points used in the fit, the effect of  $a$  scales proportionally with  $A_{kk}$  as  $N$  is varied. This is not the case for the restraint of Bayly et al.<sup>36</sup> [eq. (1)] that has  $A_{kk}$  elements given by

$$A_{kk} = \sum_{i=1}^N \frac{1}{r_{ik}^2} + \frac{a}{(q_k^2 + b^2)^{1/2}}. \quad (12)$$

The restraint term, being independent of  $N$ , would become insignificant for very large  $N$ . Consequently, the effect of their restraint is dependent on  $N$ . Two other restraining functions were tested, one linear in charge and one hyperbolic in charge, the latter similar to that of Bayly et al.<sup>36</sup> These

functions are, respectively,

$$\chi_{\text{rest}}^2 = a \sum_{j=1}^M \sum_{i=1}^N \frac{|q_j|}{r_{ij}^2} \quad (13)$$

and

$$\chi_{\text{rest}}^2 = a \sum_{j=1}^M \sum_{i=1}^N \left( (q_j^2 + b^2)^{1/2} - b \right) \frac{1}{r_{ij}^2}. \quad (14)$$

Their respective  $A_{kk}$  elements become

$$A_{kk} = \sum_{i=1}^N \frac{1}{r_{ik}^2} \left( 1 + \frac{a}{2|q_k|} \right) \quad (15)$$

and

$$A_{kk} = \sum_{i=1}^N \frac{1}{r_{ik}^2} \left( 1 + \frac{a}{2(q_k^2 + b^2)^{-1/2}} \right). \quad (16)$$

All charge fitting was performed using a modified version of the RESP module in the AMBER software package,<sup>42</sup> which takes the MEP table from GAMESS as input. Distances  $r$  were in Bohr units and charges  $q$  were in electronic charge units. Just as for the RESP method, the matrix equation was iteratively solved for the linear and hyperbolic restraining functions because their  $A_{kk}$  values depend on  $q_k$ , whereas it was solved in one step with the quadratic restraint whose  $A_{kk}$  values are independent of  $q_k$ . As shown in the Results

section, the quadratic function appeared to be the optimal function of the three to restrain charges and was therefore adopted for the remainder of the work. In addition, a number of issues were examined in comparing the quadratic function to the restraint of Bayly et al.<sup>36</sup> These issues include the dependence of the restraint on point selection and the effect of restraining the charges of hydrogen atoms.

The values of  $a$  for the quadratic restraint with the HF/6-31G\* and HF/6-31+G\* protocols are taken as 0.000184 and 0.000252, respectively. Each parameter  $a$  was calculated according to the following procedure. First the above method was used to calculate charges (with averaging; see below) for a large diverse group of 29 molecules for which there are OPLS charges, which are given in Table I and Table II. The parameter  $a$  was then varied until the slope of a plot of all unique charges versus their corresponding OPLS values equaled unity. A cross-validation analysis was then performed to test the dependence of  $a$  on the selection of molecules used in the parameterization. In this way the charges were obtained that reproduced the MEP and were comparable in size to OPLS.

Atoms that are related by symmetry are constrained to have the same charge (e.g., the hydrogens of formaldehyde). This is achieved by adding the rows for these atoms of the matrix equation [eq. (6)], reducing the dimensionality of the

**TABLE I.**  
**Slopes ( $m$ ), Correlation Coefficients ( $r$ ), and Restraint Parameters ( $a$ ) for Each Functional Group.**

Functionality	6-31G*			6-31+G*		
	$m^a$	$r^a$	$a^b$	$m^a$	$r^a$	$a^b$
Alcohol / thiol <sup>c</sup>	0.957	0.98	0.000070	0.939	0.99	0.000234
Ether / thioether <sup>d</sup>	0.777	0.94	0.000000	0.749	0.95	0.000000
Amide <sup>e</sup>	0.940	0.97	0.000144	0.857	0.97	0.000174
Aromatic <sup>f</sup>	0.957	0.98	0.000138	0.905	0.97	0.000200
Aldehyde / ketone <sup>g</sup>	1.073	0.99	0.000284	1.021	0.99	0.000408
Carboxylic acid / ester <sup>h</sup>	1.147	0.97	0.000392	1.091	0.97	0.000456
Amine <sup>i</sup>	1.019	0.98	0.000212	0.974	0.98	0.000304
Other <sup>j</sup>	1.125	0.94	0.000354	1.066	0.93	0.000450

<sup>a</sup> Derived by applying the universal  $a$  value to each group.

<sup>b</sup> Derived by fitting the charges of the molecules in each group to their OPLS counterparts to give a slope of unity.

<sup>c</sup> Water, methanol, ethanol, methanethiol, ethanethiol, phenol.

<sup>d</sup> Dimethyl ether, diethyl ether, dimethyl sulfide, diethyl sulfide.

<sup>e</sup> Formamide, acetamide, *trans*-*N*-methyl acetamide.

<sup>f</sup> Benzene, phenol, aniline, benzonitrile, chlorobenzene, benzoic acid.

<sup>g</sup> Formaldehyde, acetaldehyde, acetone.

<sup>h</sup> Acetic acid, methyl acetate, benzoic acid.

<sup>i</sup> Ammonia, methylamine, ethylamine, aniline.

<sup>j</sup> Methane, ethene, chloroethane.

**TABLE II.**  
**Listing of OPLS,<sup>13</sup> EPD / 6 – 31G\*, REPD / 6 – 31G\*, EPD / 6 – 31<sup>+</sup>G\*, and REPD / 6 – 31<sup>+</sup>G\* Charges**  
**for All 29 Molecules Used in Parameterization of a.**

Molecule	Atom	<i>q</i> (OPLS)	<i>q</i> (EPD / 6–31G*)	<i>q</i> (REPD / 6–31G*)	<i>q</i> (EPD / 6–31 <sup>+</sup> G*)	<i>q</i> (REPD / 6–31 <sup>+</sup> G*)
Methane (CH <sub>4</sub> )	C	–0.240	–0.484	–0.300	–0.484	–0.264
	H	0.060	0.121	0.075	0.121	0.066
Ethene (C <sub>2</sub> H <sub>2</sub> )	C	–0.230	–0.342	–0.328	–0.378	–0.356
	H	0.115	0.171	0.164	0.189	0.178
Water (H <sub>2</sub> O)	O	–0.834	–0.818	–0.806	–0.862	–0.844
	H	0.417	0.409	0.403	0.431	0.422
Methanol (CH <sub>3</sub> OH)	C	0.145	0.358	0.225	0.436	0.238
	O	–0.683	–0.708	–0.664	–0.765	–0.702
	HO	0.418	0.431	0.418	0.452	0.434
	HC	0.040	–0.027	0.007	–0.041	0.010
Ethanol (CH <sub>3</sub> CH <sub>2</sub> OH)	CH <sub>2</sub>	0.145	0.493	0.282	0.559	0.277
	CH <sub>3</sub>	–0.180	–0.253	–0.126	–0.259	–0.109
	OH	–0.683	–0.705	–0.648	–0.767	–0.687
	HO	0.418	0.416	0.404	0.440	0.422
	H <sub>2</sub> C	0.060	–0.076	–0.019	–0.087	–0.010
	H <sub>3</sub> C	0.060	0.067	0.042	0.067	0.039
Methanethiol (CH <sub>3</sub> SH)	C	0.060	–0.199	–0.131	–0.173	–0.102
	S	–0.435	–0.344	–0.352	–0.357	–0.364
	HS	0.255	0.210	0.207	0.215	0.211
	HC	0.040	0.111	0.092	0.105	0.085
Ethanethiol (CH <sub>3</sub> CH <sub>2</sub> SH)	CH <sub>2</sub>	0.060	0.050	0.030	0.071	0.042
	CH <sub>3</sub>	–0.180	–0.134	–0.072	–0.131	–0.059
	SH	–0.435	–0.380	–0.376	–0.393	–0.386
	HS	0.255	0.201	0.200	0.206	0.203
	H <sub>2</sub> C	0.060	0.040	0.043	0.035	0.040
	H <sub>3</sub> C	0.060	0.061	0.044	0.059	0.000
Formaldehyde (CH <sub>2</sub> O)	CO	0.450	0.444	0.390	0.475	0.405
	O	–0.450	–0.462	–0.442	–0.495	–0.471
	H	0.000	0.009	0.026	0.010	0.033
Acetaldehyde (CH <sub>3</sub> CHO)	C	0.450	0.628	0.498	0.669	0.498
	O	–0.450	–0.532	–0.496	–0.571	–0.523
	CH <sub>3</sub>	–0.180	–0.379	–0.154	–0.416	–0.127
	HC	0.000	–0.041	–0.010	–0.039	0.002
	H <sub>3</sub> C	0.060	0.108	0.054	0.119	0.050
Acetone ((CH <sub>3</sub> ) <sub>2</sub> CO)	C	0.470	0.830	0.590	0.870	0.581
	O	–0.470	–0.586	–0.536	–0.628	–0.565
	CH <sub>3</sub>	–0.180	–0.521	–0.186	–0.517	–0.128
	H	0.060	0.133	0.053	0.132	0.040
Acetic acid (CH <sub>3</sub> COOH)	C	0.520	0.886	0.711	0.927	0.706
	OH	–0.530	–0.679	–0.629	–0.708	–0.640
	OC	–0.440	–0.622	–0.569	–0.655	–0.585
	CH <sub>3</sub>	–0.180	–0.459	–0.165	–0.446	–0.106
	HO	0.450	0.454	0.445	0.468	0.454
	H <sub>3</sub> C	0.060	0.140	0.069	0.138	0.057
	C	0.510	1.006	0.771	1.060	–0.410
Methyl acetate (CH <sub>3</sub> COOCH <sub>3</sub> )	OC <sub>2</sub>	–0.330	–0.479	–0.412	–0.503	–0.599
	OC	–0.430	–0.641	–0.583	–0.676	–0.155
	CH <sub>3</sub> C	–0.180	–0.720	–0.232	–0.737	–0.029
	CH <sub>3</sub> O	0.160	0.060	–0.012	0.067	0.053
	H <sub>3</sub> CC	0.060	0.192	0.071	0.196	0.092
	H <sub>3</sub> CO	0.030	0.066	0.085	0.758	0.000

(Continued)

**TABLE II.**  
**(Continued)**

Molecule	Atom	$q(\text{OPLS})$	$q(\text{EPD} / 6-31\text{G}^*)$	$q(\text{REPD} / 6-31\text{G}^*)$	$q(\text{EPD} / 6-31^+\text{G}^*)$	$q(\text{REPD} / 6-31^+\text{G}^*)$
Ammonia ( $\text{NH}_3$ )	N	-1.026	-1.104	-1.068	-1.134	-1.080
	H	0.342	0.368	0.356	0.378	0.360
Methylamine ( $\text{CH}_3\text{NH}_2$ )	C	0.020	0.461	0.244	0.526	0.231
	N	-0.900	-1.050	-0.951	-1.113	-0.973
	HN	0.350	0.380	0.358	0.397	0.365
	HC	0.060	-0.057	-0.003	-0.069	0.004
Ethylamine ( $\text{CH}_3\text{CH}_2\text{NH}_2$ )	$\text{CH}_3$	-0.180	-0.298	-0.135	-0.280	-0.101
	$\text{CH}_2$	0.080	0.652	0.362	0.704	0.333
	N	-0.900	-1.081	-0.957	-1.149	-0.975
	$\text{H}_3\text{C}$	0.060	0.063	0.032	0.057	0.025
	$\text{H}_2\text{C}$	0.060	-0.110	-0.035	-0.121	-0.024
	$\text{H}_2\text{N}$	0.350	0.379	0.352	0.398	0.358
Formamide ( $\text{HCONH}_2$ )	C	0.500	0.741	0.461	0.802	0.433
	O	-0.500	-0.597	-0.520	-0.638	-0.537
	N	-0.760	-1.007	-0.674	-1.080	-0.643
	HC	0.000	-0.003	0.070	0.001	0.096
	HN ( <i>cis</i> )	0.380	0.450	0.351	0.479	0.350
	HN ( <i>trans</i> )	0.380	0.416	0.312	0.436	0.301
Acetamide ( $\text{CH}_3\text{CONH}_2$ )	C	0.500	0.978	0.586	1.063	0.528
	O	-0.500	-0.649	-0.555	-0.700	-0.570
	N	-0.760	-1.099	-0.745	-1.205	-0.666
	$\text{CH}_3$	-0.180	-0.559	-0.086	-0.558	-0.011
	$\text{H}_2\text{N}$ ( <i>cis</i> )	0.380	0.447	0.350	0.485	0.331
	$\text{H}_2\text{N}$ ( <i>trans</i> )	0.380	0.432	0.324	0.465	0.301
	$\text{H}_3\text{C}$	0.060	0.150	0.042	0.150	0.029
<i>trans</i> -N-Methyl acetamide ( $\text{CH}_3\text{CONHCH}_3$ )	C	0.500	0.804	0.486	0.858	0.456
	O	-0.500	-0.609	-0.535	-0.644	-0.551
	$\text{CH}_3\text{C}$	-0.180	-0.462	-0.057	-0.531	-0.028
	N	-0.500	-0.539	-0.332	-0.558	-0.306
	$\text{CH}_3\text{N}$	0.020	-0.365	-0.296	-0.406	-0.294
	$\text{H}_3\text{CC}$	0.060	0.121	0.030	0.140	0.027
	HN	0.300	0.331	0.257	0.339	0.249
	$\text{H}_3\text{CH}$	0.060	0.159	0.129	0.174	0.131
Dimethyl ether ( $(\text{CH}_3)_2\text{O}$ )	C	0.110	0.026	0.014	0.018	0.003
	O	-0.400	-0.364	-0.358	-0.378	-0.372
	H	0.030	0.052	0.055	0.057	0.061
Diethyl ether ( $(\text{CH}_3\text{CH}_2)_2\text{O}$ )	$\text{CH}_3$	-0.180	-0.315	-0.134	-0.317	-0.107
	$\text{CH}_2$	0.140	0.464	0.207	0.490	0.180
	O	-0.400	-0.508	-0.388	-0.542	-0.396
	$\text{H}_3\text{C}$	0.060	0.077	0.039	0.076	0.033
	$\text{H}_2\text{C}$	0.030	-0.063	0.002	-0.065	0.013
Dimethyl sulfide ( $(\text{CH}_3)_2\text{S}$ )	C	0.037	-0.207	-0.112	-0.185	-0.080
	S	-0.435	-0.222	-0.256	-0.236	-0.272
	H	0.060	0.106	0.080	0.101	0.072
Diethyl sulfide ( $(\text{CH}_3\text{CH}_2)_2\text{S}$ )	$\text{CH}_3$	-0.180	-0.208	-0.096	-0.188	-0.067
	$\text{CH}_2$	0.098	0.180	0.104	0.203	0.102
	S	-0.435	-0.354	-0.332	-0.374	-0.342
	$\text{H}_3\text{C}$	0.060	0.065	0.038	0.058	0.030
	$\text{H}_2\text{C}$	0.060	0.005	0.022	-0.001	0.023

(Continued)

TABLE II.  
(Continued)

Molecule	Atom	$q(\text{OPLS})$	$q(\text{EPD} / 6-31\text{G}^*)$	$q(\text{REPD} / 6-31\text{G}^*)$	$q(\text{EPD} / 6-31^+\text{G}^*)$	$q(\text{REPD} / 6-31^+\text{G}^*)$
Chloroethane (CH <sub>3</sub> CH <sub>2</sub> Cl)	CH <sub>2</sub>	-0.006	0.012	0.019	0.024	0.028
	CH <sub>3</sub>	-0.180	-0.104	-0.060	-0.100	-0.047
	Cl	-0.200	-0.231	-0.233	-0.239	-0.241
	H <sub>2</sub> C	0.103	0.085	0.080	0.084	0.079
	H <sub>3</sub> C	0.060	0.051	0.038	0.049	0.034
Benzene (C <sub>6</sub> H <sub>6</sub> )	C	-0.115	-0.133	-0.132	-0.149	-0.146
	H	0.115	0.133	0.132	0.149	0.146
Phenol (C <sub>6</sub> H <sub>5</sub> OH)	C1	0.150	0.543	0.250	0.598	0.246
	O	-0.585	-0.669	-0.582	-0.713	-0.600
	C2	-0.115	-0.387	-0.145	-0.401	-0.125
	C3	-0.115	-0.036	-0.192	-0.044	-0.206
	C4	-0.115	-0.237	-0.103	-0.242	-0.107
	H2	0.115	0.180	0.131	0.186	0.129
	H3	0.115	0.129	0.149	0.133	0.152
	H4	0.115	0.142	0.122	0.146	0.128
	HO	0.435	0.449	0.427	0.463	0.433
	C1	0.100	0.802	0.192	0.959	0.181
Aniline (C <sub>6</sub> H <sub>5</sub> NH <sub>2</sub> )	N	-0.900	-1.263	-0.766	-1.439	-0.776
	C2	-0.115	-0.478	-0.137	-0.528	-0.117
	C3	-0.115	-0.016	-0.192	-0.014	-0.203
	C4	-0.115	-0.292	-0.107	-0.312	-0.111
	H2	0.115	0.182	0.122	0.192	0.122
	H3	0.115	0.130	0.142	0.136	0.143
	H4	0.115	0.149	0.115	0.156	0.118
	H <sub>2</sub> N	0.400	0.484	0.348	0.532	0.349
Chlorobenzene (C <sub>6</sub> H <sub>5</sub> Cl)	C1	0.180	0.008	0.016	0.044	0.037
	Cl	-0.180	-0.139	-0.136	-0.150	-0.142
	C2	-0.115	-0.023	-0.049	-0.033	-0.051
	C3	-0.115	-0.205	-0.156	-0.213	-0.162
	C4	-0.115	-0.070	-0.118	-0.073	-0.124
	H2	0.115	0.115	0.119	0.117	0.119
	H3	0.115	0.151	0.138	0.155	0.140
	H4	0.115	0.125	0.134	0.127	0.137
Benzonitrile (C <sub>6</sub> H <sub>5</sub> CN)	C1	0.035	0.101	0.067	0.146	0.088
	CN	0.395	0.302	0.289	0.297	0.287
	C2	-0.115	-0.171	-0.128	-0.190	-0.128
	C3	-0.115	-0.133	-0.155	-0.139	-0.166
	C4	-0.115	-0.081	-0.072	-0.077	-0.070
	H2	0.115	0.167	0.157	0.173	0.159
	H3	0.115	0.140	0.141	0.143	0.144
	H4	0.115	0.130	0.132	0.130	0.134
	N	-0.430	-0.458	-0.446	-0.470	-0.457
	C1	-0.115	0.047	0.077	0.121	0.121
Benzoic acid (C <sub>6</sub> H <sub>5</sub> COOH)	CO <sub>2</sub>	0.635	0.710	0.564	0.697	0.524
	C2	-0.115	-0.157	-0.122	-0.188	-0.119
	C3	-0.115	-0.146	-0.157	-0.155	-0.173
	C4	-0.115	-0.083	-0.105	-0.073	-0.103
	H2	0.115	0.155	0.129	0.166	0.126
	H3	0.115	0.137	0.142	0.142	0.148
	H4	0.115	0.125	0.136	0.123	0.138
	OC	-0.440	-0.587	-0.527	-0.603	-0.528
	OH	-0.530	-0.664	-0.565	-0.684	-0.551
	HO	0.450	0.474	0.436	0.489	0.435



matrix.<sup>36</sup> However, atoms that are not symmetry related but conformationally equivalent should have their charges constrained to be identical if conformational flexibility is allowed in a simulation. This issue of averaging conformationally equivalent atoms is merely a special case of general conformational dependence in which the new conformation is identical to the old. This work uses a two-stage averaging procedure similar to that of Bayly et al.<sup>36</sup> A charge fit as described above is first performed without averaging. Then, while freezing the charges of all atoms not requiring averaging nor adjacent to ones that do, a second fit is performed with conformationally equivalent atoms constrained to be the same. This freezing in the second stage prevents the averaging from detrimentally affecting the charges of the rest of the molecule. Unlike Bayly et al.,<sup>36</sup> the same restraint  $a$  is applied in both stages. An alternative method of simply averaging conformationally equivalent charges after an unaveraged fit was discarded on the grounds that the final charges are then not optimized to reproduce the MEP. However, one unavoidable consequence for any averaging method is that the quality of the fit is reduced. The amide hydrogens were excluded from the conformational averaging because their barrier to rotation is sufficiently high to keep them distinguishable.

Four sets of charges were calculated. They differed in their basis set (6-31G\* or 6-31+G\*) and whether a restraining function was used. Each set was labeled as EPD/6-31G\*, REPD/6-31G\*, EPD/6-31+G\*, or REPD/6-31+G\*.

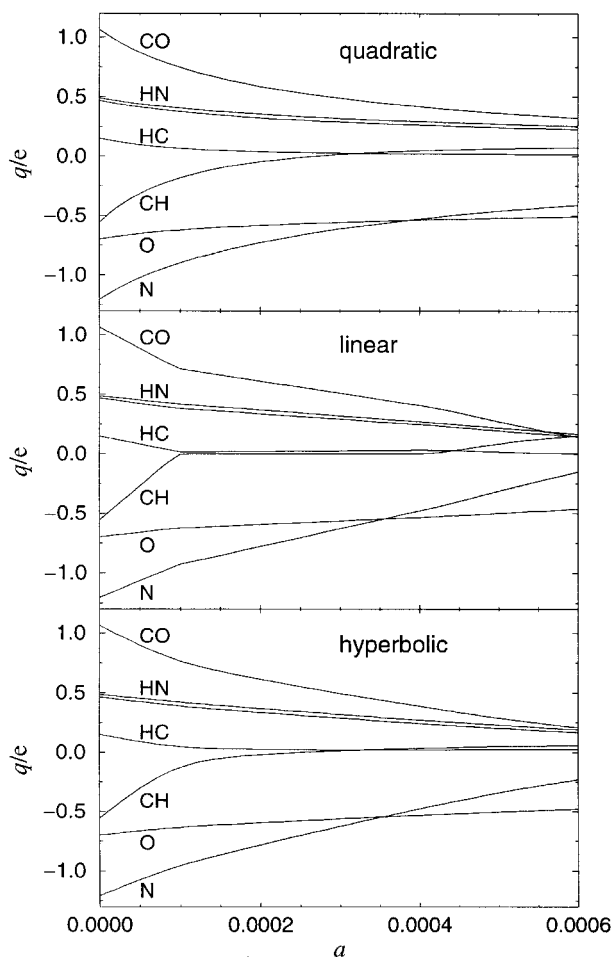
One final issue examined was the conformational dependence of charges. The charges produced for each of the five conformations of 1-propylamine were applied to the other four conformations. The dipole moment and quality of fit of the resulting CEP were compared to the values from the original *ab initio* calculations.

## Results

Aspects involving the development of the method include the choice of restraining function, the relationship between the restraint and point density and choice of the point density itself, the effect of restraining hydrogen atoms, the derivation of  $a$ , the suitability of linear charge scaling, and the cross-validation analysis for the parameter  $a$ . In the following discussion 6-31+G\* acetamide

is used as a typical molecule to demonstrate each point.

Figure 1 shows how the charges for acetamide vary with  $a$  for each of the three restraining functions: quadratic [eq. (10)], linear [eq. (13)], and hyperbolic [eq. (14)]. The value of  $b$  for the hyperbolic restraint is taken as 0.1, the same as for RESP. The one feature that all restraints have in common is that they restrain buried atoms the most. The linear restraint applies a constant force to each charge. When one charge reaches zero it generally stays there, although it may deviate again slightly to preserve the overall molecular charge. At these points, all other charges suffer discontinuities in their slopes. The linear restraint is unsatisfactory because some charges reach zero before other charges are even mildly affected. Setting buried charges to zero or some other small value may be a promising way of developing transferable force

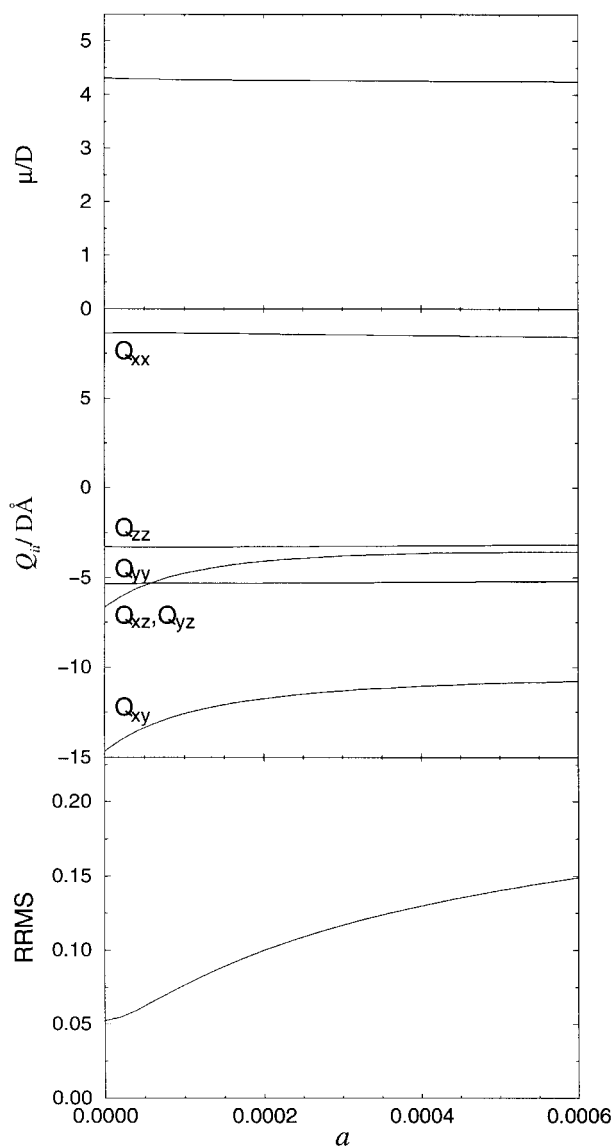


**FIGURE 1.** Effect of the restraint  $a$  on the 6-31+G\* charges for acetamide for the quadratic, linear, and hyperbolic restraining functions.

fields; but because the overall goal is to make charges reproduce OPLS parameters, this restraint was not pursued. Using a hyperbolic restraint produces similar results but with these gradient discontinuities rounded off. This, together with the need for the extra parameter  $b$ , ruled out the use of a hyperbolic function. The quadratic restraint applies a restraining force proportional to the charge, thus restraining larger charges more. Large charges are generally found for polar atoms or buried atoms. Bayly et al.<sup>36</sup> decided against using a quadratic restraint because larger, polar charges experience a stronger restraint than smaller, non-polar charges. However, it is rather the overall proximity of the atom to the fitting points that determines the sensitivity of its charge to a restraint. As Figure 1 indicates, the charges of the buried atoms (the carbons and nitrogen) are rapidly reduced in magnitude while the well-exposed polar hydrogens and oxygen charges change much more slowly. Figure 2 illustrates how the dipole moment, quadrupole moment, and RRMS vary with the quadratic restraint. The dipole moment and diagonal elements of the quadrupole moment are hardly affected by the restraint, while the off-diagonal elements of the quadrupole moment decrease in magnitude to some extent. A moderate increase in the RRMS accompanies the restraint, but the restrained charges still adequately reproduce the MEP.

Charges were found to converge quite slowly with point density. While it was suggested that thousands of points are necessary to obtain well-defined charges,<sup>10</sup> the singular value decomposition analysis of Francel et al.<sup>28</sup> suggests that the MEP does not contain sufficient information to produce such charges, regardless of the number of points. However, the quality of the fit to the MEP is still adequate at point densities of 1 point  $\text{\AA}^{-2}$ , the density used in this work. This is demonstrated in Figure 3 that shows the variation of the RESP/6-31+G\* charges for acetamide with the point density used to calculate the charges. At densities above 1 point  $\text{\AA}^{-2}$ , the charge variation is insignificant for the purposes of reproducing the MEP. The RRMS at this density is calculated to be 0.109 and barely changes to 0.108 at a density of 100 points  $\text{\AA}^{-2}$ .

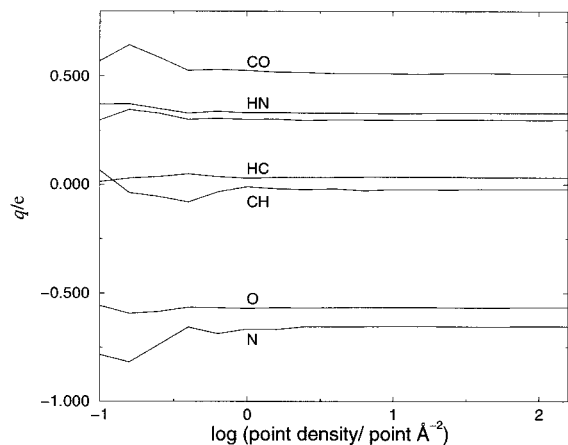
The effect of the restraining function should be independent of the number of points used. This is not only convenient if different point densities are desired, but it is also essential if the restraining function is to act in a similar way on different molecules with different numbers of points. Figure



**FIGURE 2.** Effect of the quadratic restraint parameter  $a$  on the dipole moment, quadrupole moment, and RRMS for 6-31+G\* acetamide.

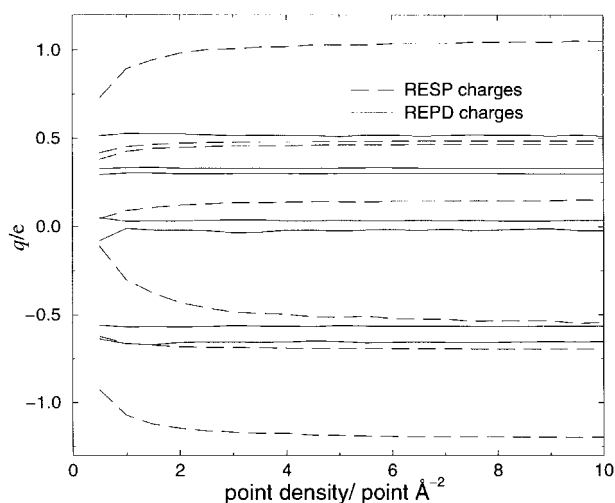
4 shows how RESP/6-31+G and RESPD/6-31+G\* charges for acetamide vary with point density. Clearly, the effect of the additive RESP restraining function is diminished at higher point densities while the multiplicative RESPD restraining function acts on all charges similarly for all point densities.

A final note about point selection is that care must be taken to not include points closer than approximately 1.4 times the molecule's van der Waals surface. First, it is unnecessary to include such close points because the atoms of other molecules are unable to approach to this distance and, second, it may affect the values of charges

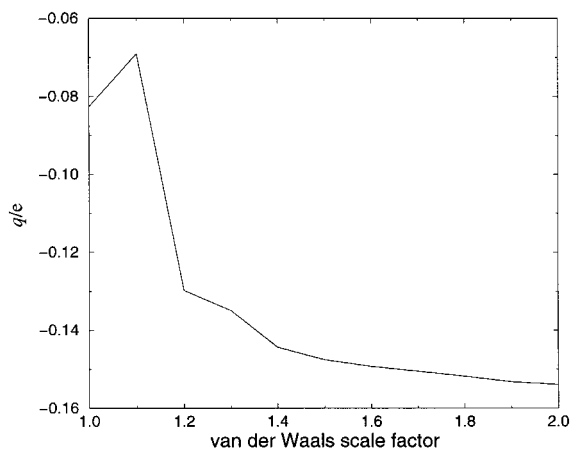


**FIGURE 3.** Variation of the charges with point density for REPD/6-31+G\* acetamide.

obtained. To illustrate this point, charges are calculated using a single sphere of fitting points of different radii. Figure 5 reveals how the carbon charge of 6-31+G\* benzene is severely influenced by the distance at which the sphere of fitting points is chosen. The carbon charge becomes much smaller in magnitude at shorter distances. This effect is probably due to a severe worsening of the point charge approximation and such charges, which are dependent on the radius chosen, are questionable, as well as being unsuitable for modeling electrostatic interactions at longer ranges. The CHELPG method<sup>10</sup> has been frequently implemented (GAMESS,<sup>39</sup> Gaussian 94<sup>37</sup>), including these close points.



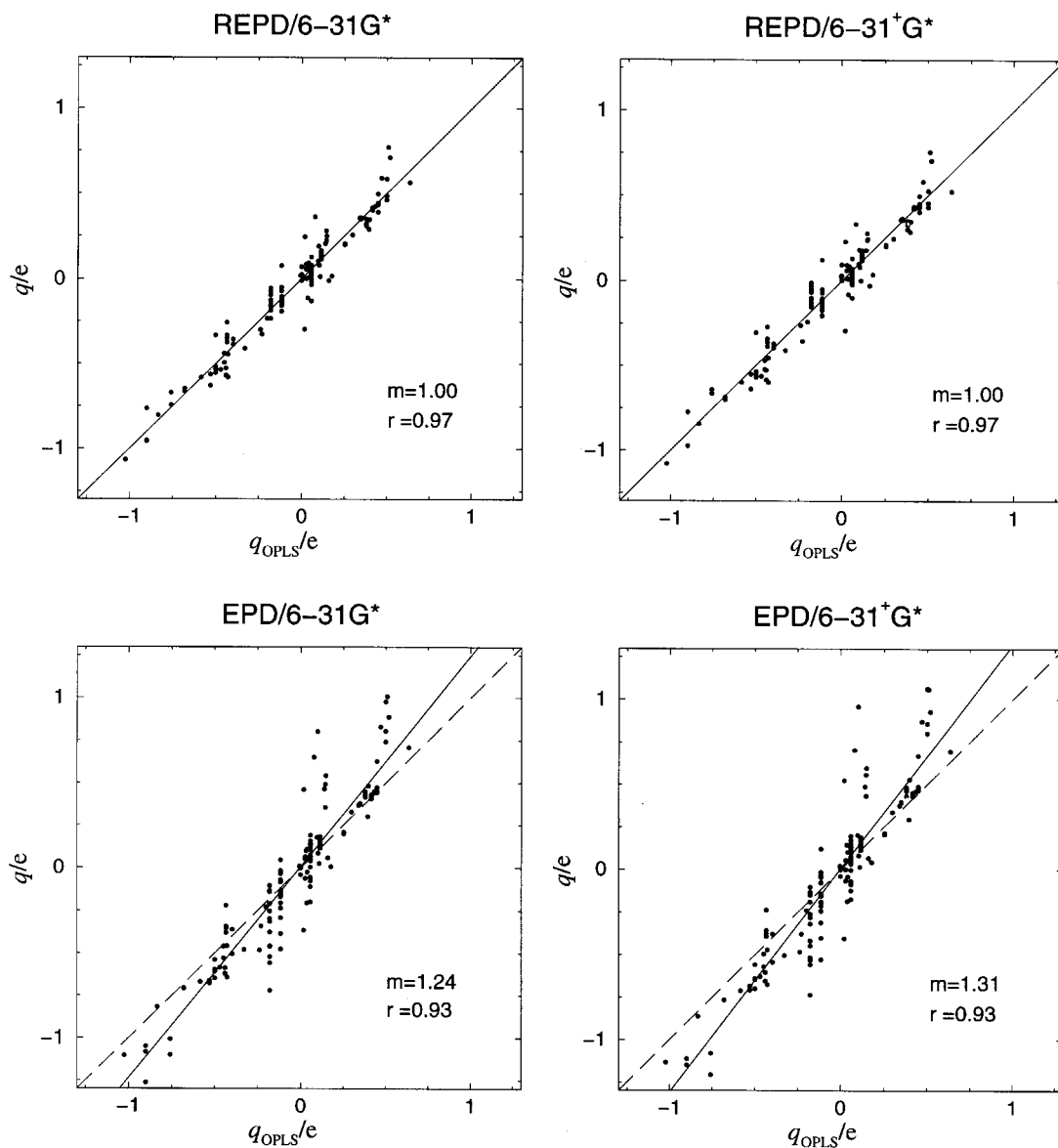
**FIGURE 4.** Comparison between (---) RESP/6-31+G\* and (—) REPD/6-31+G\* charges with  $a = 0.000252$  for acetamide as the point density is varied.



**FIGURE 5.** Influence of the radius of the fitting point sphere (determined by the van der Waals scale factor) on the carbon charge calculated for EPD/6-31+G\* benzene.

The restraint in this work was applied to the charges of all atoms without exception. Bayly et al.<sup>36</sup> chose not to restrain hydrogen atom charges because they are already well defined and therefore do not require restraining. However, we feel that this exclusion is inconsistent and unnecessary. Atom type does not necessarily correlate with the need for a restraint, because on the one hand it could be argued that other well-defined atoms such as carbonyl oxygens might also not require restraining and on the other hand, hydrogen atoms in large, folded molecules might be sufficiently buried as to require restraining. In either case, whether or not hydrogens are restrained makes little difference to the charges for two reasons. First, there is strong coupling between adjacent atoms, which is especially significant for atoms with only one neighbor such as hydrogens. Thus, any change in the charge of an atom neighboring a hydrogen will affect the hydrogen charge indirectly. Second, the charges of all well-defined surface atoms, usually including hydrogens, are relatively insensitive to the restraint as is evident in Figure 1. To verify this, charges calculated as a function of the restraint with and without the hydrogens restrained were found to differ negligibly (data not shown). Therefore, for the reasons described above, the charges of *all* atoms were chosen to be restrained.

As previously mentioned, the  $a$  parameters that optimize the REPD charges to OPLS charges are 0.000184 and 0.000252 for the 6-31G\* and 6-31+G\* basis sets, respectively. Figure 6 shows correlation plots of the set of unique charges calculated for



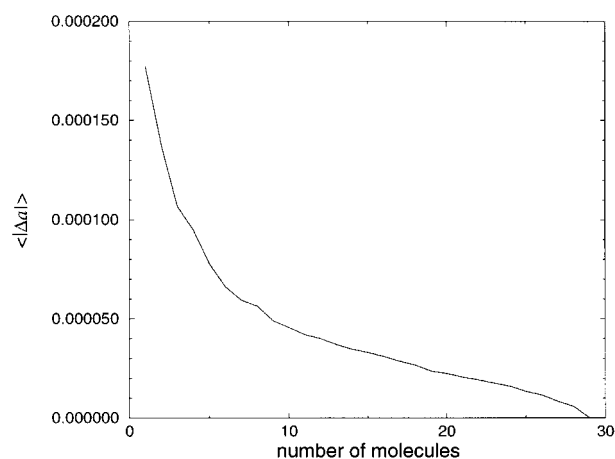
**FIGURE 6.** Correlation plots of REPD/6-31G\*, REPD/6-31<sup>+</sup>G\*, EPD/6-31G\*, and EPD/6-31<sup>+</sup>G\* charges versus the corresponding OPLS parameters for the 29 molecules given in Table II.

each basis set versus their OPLS counterparts. For both sets of REPD charges, the slopes of the lines are, of course, 1.00 and there is a good correlation with OPLS charges with identical correlation coefficients of 0.97. The correlation is poorer for unrestrained charges with a correlation coefficient of 0.93 for EPD/6-31G\* and EPD/6-31<sup>+</sup>G\* charges and slopes of 1.24 and 1.31, respectively. In addition, a correlation plot of one restrained charge set against the other (not shown) gives a correlation coefficient of 1.00 and the expected slope of unity. An alternative method of adjusting charges is the use of linear scaling as is commonly used for

charges fitted to semiempirical MEPS.<sup>43</sup> Such a technique can be used to produce charges that have a slope of unity versus their respective OPLS values. For example, 0.761 is the appropriate scaling factor for 6-31<sup>+</sup>G\* charges. However, linear scaling does not alter the correlation coefficient that remains at 0.93 in this case. Furthermore, the quality of the fit of the CEP with the MEP is consequently reduced. For scaled 6-31<sup>+</sup>G\* acetamide, for example, the RRMS jumps significantly to 0.245 compared to 0.109 for REPD/6-31<sup>+</sup>G\* acetamide. Also, the dipole moment, as would be expected, undesirably scales

down by the same factor. The results are similar for 6-31G\* charges for which the appropriate scale factor is 0.807. Therefore, use of a restraint in the fitting procedure is clearly superior to a simple linear scaling of EPD charges.

To determine the dependence of the parameter  $a$  on the molecules used in the parameterization, a cross-validation analysis was used. In this procedure, the 29 molecules were divided up randomly into five groups, each containing five or six molecules. For each group the  $a$  parameter was derived from the remaining four groups and then applied to the original group to predict charges for the selection of molecules in that group. A correlation coefficient between these charges and their OPLS counterparts was then calculated. This process was repeated 100 times and gave 500 different selections of the 29 molecules. The average correlation coefficients for the predicted charges with OPLS charges were found to be 0.97 for REPD/6-31G\* and REPD/6-31<sup>+</sup>G\*, the same as the overall correlation coefficient. To provide further insight into how  $a$  depends on the choice of molecules, Figure 7 shows how the absolute deviation of  $a$  from its universal value varies with the number of molecules used to derive  $a$  for the 6-31<sup>+</sup>G\* basis set. This graph was obtained by averaging over 300 random molecule sequences and, by definition, reached zero at 29 molecules. While there was a clear decreasing dependence of  $a$  on molecule selection as more molecules were used, this dependence was still nonnegligible for 20 molecules and had a variation of around



**FIGURE 7.** Variation of the absolute error in  $a$  from its universal value with the number of molecules used to derive  $a$  averaged over 300 random molecule sequences.

$\pm 0.000020$ . This corresponded to an average charge variation of  $\pm 0.01$ . Clearly, no single  $a$  parameter is perfectly suited to all molecule types. To test how the  $a$  parameter performs for each chemical functionality, the molecules were divided into groups chosen according to functionality (Table I). Some aromatic molecules are put into two groups, because they contain two of the listed functionalities. First the universal parameter was applied to each functionality, giving a slope and correlation coefficient. Then a specific  $a$  parameter was calculated using only the molecules in each group. The closeness of the correlation coefficient and the gradient to unity and the individual  $a$  values to the universal  $a$  value indicates the appropriateness of the general restraint for that functionality. Table I reveals that the ethers, sulfides, and amides are overrestrained by the universal  $a$  value. The zero value of  $a$  for the ethers and sulfides arises because their unrestrained charges are actually smaller than OPLS charges. On the other hand, the carbonyl compounds are not sufficiently restrained. While it is conceivable that different  $a$  parameters could be calculated for different functional groups, it would be impossible to calculate  $a$  values for functionalities not covered by the OPLS force field; thus, the method only remains practicable if the universal  $a$  value is used. It is nevertheless indisputable that the REPD procedure yields charges much closer to their OPLS counterparts than the EPD method.

The charges obtained using all four protocols together with the OPLS charges<sup>13</sup> are given in Table II for all 29 molecules. There are a number of features to note. As expected, EPD charges are larger in magnitude than REPD charges and OPLS charges, especially for buried atoms, while the charges of surface atoms such as in hydroxy groups decrease only a small amount on restraint. EPD/6-31<sup>+</sup>G\* charges are generally larger in magnitude than EPD/6-31G\* charges while REPD/6-31<sup>+</sup>G\* and REPD/6-31G\* charges are comparable in magnitude with respect to each other and to OPLS charges, as indeed they are designed to be. Thus, as mentioned above, a larger value of  $a$  is needed to restrain the larger EPD/6-31<sup>+</sup>G\* charges to their corresponding OPLS values. In our accompanying article<sup>44</sup> further comparisons are made between all four charge sets in relation to their dipole moments and free energies of hydration.

One final aspect of REPD charges to be discussed is their conformational dependence. This

may be assessed by observing how the dipole moment and RRMS of 1-propylamine vary when the charges derived from one conformation are applied to the other conformations.<sup>8,9</sup> Ideally, this variation should be negligible. The results are shown in Table III. For REPD charges the dipole moments generally deviate less from *ab initio* and the increase in RRMS is smaller than for EPD charges. It can therefore be concluded that the REPD charges are less conformationally dependent and better able to reproduce the MEP for other conformations. In the particular case where charges are applied to their own conformation, however, the relative performance of EPD and REPD charges is less conclusive and REPD is better for three conformations and EPD is better for the other two. The *g* – *g* + REPD charge sets require particular attention, because the errors in dipole moment and RRMS are particularly small for all conformations, suggesting that this charge set is the most suitable for flexible 1-propylamine simulations. Coinciden-

tally, this is also the charge set with the smallest charge on the nitrogen. These observations regarding dipole moment and RRMS are mainly due to the fact that restrained charges are smaller in magnitude and so the variation of charges with conformation will be correspondingly smaller. One method of dealing with conformational dependence is to produce an averaged set of charges. Reynolds et al.<sup>8</sup> performed this averaging over different conformations, and each conformation was Boltzmann weighted according to its energy. However, the method is more computationally demanding, only conformational energy minima are considered, and the charges produced are temperature dependent. The main problem remains that there is generally no unique charge set able to reproduce the MEP well for all the conformations of a molecule. Charges that actually vary with conformation may be implemented in the force field,<sup>14</sup> although as yet this solution has not proved to be popular.

**TABLE III.**  
**Comparison of Conformational Dependence between EPD and REPD Charges in Relation to Dipole Moments and RRMS of Fit for 1-Propylamine.**

Effect on Dipole Moment											
Test Conformation <sup>a</sup>	<i>Ab Initio</i>	6–31+G* Charge Set									
		<i>aa</i>		<i>ag</i>		<i>ga</i>		<i>g</i> – <i>g</i> +		<i>g</i> + <i>g</i> +	
		EPD	REPD	EPD	REPD	EPD	REPD	EPD	REPD	EPD	REPD
<i>aa</i>	1.50	<b>1.45<sup>b</sup></b>	<b>1.39</b>	1.52	1.47	2.24	1.88	1.88	1.54	2.26	1.87
<i>ag</i>	1.57	1.40	1.35	<b>1.49</b>	<b>1.45</b>	2.16	1.82	1.94	1.54	2.14	1.81
<i>ga</i>	1.44	1.99	1.76	2.05	1.80	<b>1.77</b>	<b>1.60</b>	1.86	1.60	1.67	1.54
<i>g</i> – <i>g</i> +	1.50	2.27	2.10	2.09	1.84	2.37	2.16	<b>1.55</b>	<b>1.53</b>	2.00	1.74
<i>g</i> + <i>g</i> +	1.41	2.01	1.79	2.05	1.80	1.80	1.64	1.89	1.61	<b>1.60</b>	<b>1.52</b>
Sum of errors <sup>c</sup>		2.14	1.63	1.94	1.24	2.92	1.68	1.70	0.46	2.25	1.06
6–31G* Charge Set <sup>d</sup>											
Sum of errors <sup>c</sup>		2.11	1.09	2.05	1.14	2.09	1.08	1.56	0.44	1.38	0.60
Effect on RRMS											
Test Conformation <sup>a</sup>	<i>Ab Initio</i>	6–31+G* Charge Set									
		<i>aa</i>		<i>ag</i>		<i>ga</i>		<i>g</i> – <i>g</i> +		<i>g</i> + <i>g</i> +	
		EPD	REPD	EPD	REPD	EPD	REPD	EPD	REPD	EPD	REPD
		<b>0.21<sup>b</sup></b>	<b>0.22</b>	0.22	0.23	0.43	0.31	0.32	0.28	0.46	0.33
		0.27	0.30	<b>0.25</b>	<b>0.26</b>	0.43	0.36	0.36	0.31	0.41	0.33
		0.37	0.31	0.39	0.33	<b>0.28</b>	<b>0.26</b>	0.34	0.31	0.32	0.29
		0.46	0.41	0.42	0.34	0.54	0.45	<b>0.30</b>	<b>0.30</b>	0.42	0.35
		0.44	0.38	0.42	0.34	0.38	0.36	0.37	0.31	<b>0.29</b>	<b>0.28</b>
Total RRMS		1.75	1.62	1.70	1.50	2.06	1.74	1.69	1.51	1.90	1.58
6–31G* Charge Set <sup>d</sup>											
Total RRMS		1.76	1.54	1.66	1.42	1.75	1.53	1.59	1.43	1.59	1.39

<sup>a</sup> Conformations described by (lp)-N-C-C and N-C-C-C torsions. Anti and gauche conformations are denoted *a* and *g*, respectively.  
<sup>b</sup> Bold numbers indicate that charges are tested on the same conformation from which they were derived.  
<sup>c</sup> The error is taken relative to the *ab initio* dipole moment.  
<sup>d</sup> Only the errors and total RRMS are shown. The full table is similar to the 6–31+G\* results.

## Conclusion

A method based on the RESP procedure of Bayly et al.<sup>36</sup> was proposed that, by use of a restraining function, produces atomic charges of comparable size to the equivalent OPLS charges for almost all atoms. The restraint used is quadratic in charge, produces charges independent of point density, restrains all charges including hydrogens, and only requires a single parameter  $a$ . The REPD charges calculated at the 6–31G\* and 6–31+G\* levels are able to reproduce their OPLS counterparts with a correlation coefficient of 0.97, which is an improvement on the value 0.93 for charges produced without the restraint. A cross-validation analysis demonstrates the insensitivity of the calculated charges to the choice of molecules used in the parameterization. However, a more detailed analysis based on functionality reveals that there is still some small but inevitable dependence of  $a$  on the molecular selection. Of course, it would still be possible to parameterize  $a$  to reproduce some other charge set or property such as free energies of hydration if so desired. Although the proposed method generates charges that are comparable in size to those of the OPLS force field, the raw OPLS parameters are specifically designed to reproduce condensed phase properties and in our opinion should therefore (when available) be the parameters of choice in condensed phase simulations. The REPD method, however, is suitable when parameterization based on condensed phase properties is not possible. This is especially the case for large molecules where standard OPLS parameters are available for some functionalities but not others. In our accompanying article<sup>44</sup> the free energies of hydration are calculated for a large number of molecules to determine the reliability of REPD charges in condensed phase computer simulations.

## Acknowledgments

We would like to thank the Commonwealth Scholarship Commission for financial support in the form of a scholarship for R.H.H. and Prof. W. L. Jorgensen for making available the most recent OPLS all-atom parameter set. J.W.E. is a Royal Society University Research Fellow.

## References

- Williams, D. E. *J Comput Chem* 1994, 15, 719.
- Sokalski, W. A.; Keller, D. A.; Ornstein, R. L.; Rein, R. *J Comput Chem* 1993, 14, 970.
- Angyan, J. G.; Chipot, C. *Int J Quantum Chem* 1994, 52, 17.
- Wallqvist, A.; Berne, B. J. *J Phys Chem* 1993, 97, 13841.
- Williams, D. E. *Biopolymers* 1990, 29, 1367.
- Stouch, T. R.; Williams, D. E. *J Comput Chem* 1992, 13, 622.
- Stouch, T. R.; Williams, D. E. *J Comput Chem* 1993, 14, 858.
- Reynolds, C. A.; Essex, J. W.; Richards, W. G. *J Am Chem Soc* 1992, 114, 9075.
- Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. *J Am Chem Soc* 1993, 115, 9620.
- Breneman, C. M.; Wiberg, K. B. *J Comput Chem* 1990, 11, 361.
- Colonna, F.; Evleth, E. M. *Chem Phys Lett* 1993, 212, 665.
- Su, Z. W.; Coppens, P. *Naturforsch* 1993, 48a, 85.
- Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J Am Chem Soc* 1996, 118, 11225.
- Rappe, A. K.; Goddard, W. A. *J Phys Chem* 1991, 95, 3358.
- King, W. T.; Mast, G. B. *J Phys Chem* 1976, 80, 2521.
- Newton, J. H.; Person, W. B. *J Chem Phys* 1976, 64, 3036.
- Cioslowski, J. *J Am Chem Soc* 1989, 111, 8333.
- Mulliken, R. S. *J Chem Phys* 1955, 23, 1833.
- Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J Chem Phys* 1985, 83, 735.
- Hirschfeld, F. L. *Theor Chim Acta* 1977, 44, 129.
- Bader, R. F. W. *Chem Rev* 1991, 91, 893.
- Cox, S. R.; Williams, D. E. *J Comput Chem* 1981, 2, 304.
- Wiberg, K. B.; Rablen, P. R. *J Comput Chem* 1993, 14, 1504.
- Duffy, E. M.; Severance, D. L.; Jorgensen, W. L. *Isr J Chem* 1993, 33, 323.
- Momany, F. A. *J Phys Chem* 1978, 82, 592.
- Singh, U. C.; Kollman, P. A. *J Comput Chem* 1984, 5, 129.
- Chirlian, L. E.; Francl, M. M. *J Comput Chem* 1987, 8, 894.
- Francl, M. M.; Carey, C.; Chirlian, L. E.; Gange, D. M. *J Comput Chem* 1996, 17, 367.
- Soliva, R.; Orozco, M.; Luque, F. J. *J Comput Chem* 1997, 18, 980.
- Chipot, C.; Maigret, B.; Rivail, J. L.; Scheraga, H. A. *J Phys Chem* 1992, 96, 10276.
- Stone, A. J.; Alderton, M. *Mol Phys* 1985, 56, 1047.
- Colonna, F.; Evleth, E.; Angyan, J. G. *J Comput Chem* 1992, 13, 1234.
- Ferency, G. G. *J Comput Chem* 1991, 12, 913.
- Kuyper, L. F.; Hunter, R. N.; Ashton, D.; Merz, K. M.; Kollman, P. A. *J Phys Chem* 1991, 95, 6661.
- Cao, M.; Teppen, B. J.; Miller, D. M.; Pranata, J.; Schafer, L. *J Phys Chem* 1994, 98, 11353.
- Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J Phys Chem* 1993, 97, 10269.

37. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, G. A. P. T.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. GAUSSIAN 94; Gaussian Inc.: Pittsburgh, PA, 1995; Revision d.4 edition.
38. Connolly, M. L. *J Appl Crystallog* 1983, 16, 548.
39. Schmidt, M. W.; Baldrige, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J Comput Chem* 1993, 14, 1347.
40. Woods, R. J.; Khalil, M.; Pell, W.; Moffat, S. H.; Smith, V. H. *J Comput Chem* 1990, 11, 297.
41. Spackman, M. A. *J Comput Chem* 1996, 17, 1.
42. Pearlman, D. A.; Caldwell, J. W.; Seibel, G.; Weiner, P. A.; Kollman, P. A. AMBER 4.1 (UCSF); Department of Pharmaceutical Chemistry, University of California: San Francisco, CA, 1991.
43. Merz, K. M. *J Comput Chem* 1992, 13, 749.
44. Henchman, R. H.; Essex, J. W. *J Comput Chem* 1999, 20, 499.